DOCUMENT RESUME

ED 426 662 HE 031 818

AUTHOR	Bebeau, Muriel J.; Thoma, Stephen J.
TITLE	Designing and Testing a Measure of Intermediate Level Ethical Concepts.
PUB DATE	1998-04-00
NOTE	19p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Case Studies; Cognitive Development; Concept Formation; Construct Validity; Context Effect; Dental Students; *Ethical Instruction; *Ethics; Generalization; Higher Education; Individual Development; *Moral Development; Predictive Measurement; Predictive Validity; Professional Education; Student Development; *Test Validity; Values

ABSTRACT

This study describes the steps taken to measure the integrity of an intermediate concept measure and test the assessment strategy of a well-validated dental ethics curriculum. Because this curriculum, which had been in place for 15 years, had consistently been shown to influence moral judgment, it was posited that a cross-sectional comparison between senior dental students and entering freshmen would show the educational effects of the ethics curriculum on intermediate concepts and make it possible to examine whether the acquisition of more general philosophical ideals enhances a student's ability to select appropriate actions and justifications in unfamiliar situations. Using a five-step process, the study presented five cases to 12 dental residents and four faculty, and their action choices and justifications were ranked. These ranked responses were then submitted to a second group consisting of two faculty and three advanced residents. In a final step, 18 teachers of dental ethics were asked to respond to the cases. The findings supported the hypothesis that courses that present intermediate ethical concepts enhance the ability to resolve familiar problems, but transfer to unfamiliar problems depends on whether students understand more advanced conceptual frameworks. Four tables present case data; one sample case is appended. (Contains 28 references.) (CH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 426 662

Designing and Testing a Measure of Intermediate Level Ethical Concepts

Muriel J. Bebeau and Stephen J. Thoma
University of Minnesota and University of Alabama

Paper Presented at the
Annual Meeting of the American Educational Research Association
San Diego, CA, April 13-17, 1998

Abstract

Professionals with a good grasp of abstract moral schemas may nonetheless lack an adequate grasp of specific intermediate concepts (e.g., informed consent, confidentiality, paternalism, etc.) that apply to their profession (Strike, 1982). Typically, ethics courses focus on more concrete intermediate concepts and codes rather than abstract moral schemas, but educators wishing to assess outcomes must rely on tests of life-span development (e.g., Rest's Defining Issues Test [DIT]), as profession-specific measures of ethical reasoning have not been validated. This study describes steps taken to insure the integrity of an intermediate concepts measure (ICM) and tests the assessment strategy in the context of a well-validated dental ethics curriculum (Bebeau, 1994), reasoning that if the strategy is judged useful in one professional context, it would have application elsewhere. Construct validity is demonstrated by (1) high levels of agreement among 14 dental ethicists as to better and worse action choices and justifications for ethical problems presented on the ICM; (2) significant differences among groups expected to differ in ethics interest and expertise (68 college freshmen vs. 86 dental freshmen vs. 74 dental seniors); (3) moderate correlations between the ICM and DIT; nonsignificant correlations between the ICM and GPA; (4) educational intervention effects (seniors completing the ICM and the DIT as part of a final curriculum assessment scored significantly higher than freshmen, who provided baseline data); and (5) stronger relationships between justifications and action for the "more expert" group. Findings support the hypothesis that courses that present intermediate ethical concepts enhance the ability to resolve familiar problems, but transfer to unfamiliar problems depends on whether students have access to the more advanced conceptual frameworks measured by the DIT.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Muriel J. Bebeau

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Reprint requests should be addressed to: Muriel J. Bebeau, Department of Preventive Sciences, School of Dentistry, University of Minnesota, 515 Delaware St. S.E., Minneapolis, MN 55455; Phone: (612) 625-4633; Fax: (612) 626-6096; E-mail: bebea001@maroon.tc.umn.edu

AE 031 218

BEST COPY AVAILABLE

Designing and Testing a Measure of Intermediate Level Ethical Concepts

This paper describes initial efforts to design and validate a method of assessment that could be used as an outcome measure for professional ethics instruction. In the first section, we describe problems encountered by educators attempting to assess the outcomes of ethics education, the theoretical grounding and rationale for the design of a measure of intermediate level ethical concepts, and specific measurement problems to overcome when designing and validating a measure of intermediate concepts. In the methods section, a strategy for ensuring the integrity of the intermediate concept measure is described. We test the measurement strategy in the context of a well-validated ethics curriculum, reasoning that if the method is judged useful for our context--dental ethics education--it would have application in other professional education settings. To build the case for construct validity, groups expected to differ in moral judgment development and dental ethics expertise are tested with the new measure. To explore the role of ethics education and moral judgment development on acquisition of intermediate ethical concepts and transfer to novel problems, we selected cases that were familiar and unfamiliar to our "more expert" dental student group and compared their responses to novice groups.

Theoretical Foundations

Preparing professionals to discern the right course of action in problematic work contexts has become an immense enterprise. Nearly every professional school in the country (Thornton, Callahan, & Nelson, 1993) now includes some form of ethics instruction as a regular part of the curriculum. Despite the energy expended in designing ethics courses, only a handful of studies (see Rest & Narvaez, 1994 for a summary of outcome studies in the professions) describe the effects of instruction on highly publicized goals for professional ethical development (Hastings Center, 1980). In dentistry, for example, ethics courses were introduced in the early 80s and are now a required part of the curriculum (American Dental Association, 1991). Yet, just two studies (Bebeau, 1994; Bebeau & Thoma, 1994) describe the effects of instruction on the ethical decision making competencies described in American Association of Dental Schools (1989) curriculum guidelines. Earlier studies in dentistry (Parrish, 1968; Morris & Sherlock, 1971; Steinberg, 1973 Moody, Van Tassel, & Cash, 1974; Lancaster, 1989;), as in other professions, describe attitude changes presumed to be linked to values that influence ethical decision making.

One problem for those who wish to study the effects of instruction is the lack of well-validated outcome measures. Researchers must rely on (1) instructor-designed tests that often lack sufficient evidence of validity and reliability, (2) on attitude indicators that, though well-validated, may not predict to the specific ethical competencies taught in the course, or (3) on measures of life-span moral judgment development that, although well-validated, may not be sensitive to the particular goals of an ethics course. Researchers who have reflected on the use of measures of life-span development, e.g., The Defining Issues Test (Rest, 1979), The Moral Judgment Interview (Colby & Kohlberg, 1987), or The Sociomoral Reflection Measure (Gibbs, Basinger & Fuller, 1992) cite several issues related to instrument use and measurement design.

First, tests of moral judgment development are based on Kohlbergian theory that describes highly abstract markers of life-span development. According to the theory (and research), stages function as default schemas, providing conceptual guidance for action choice in situations that present conflicting moral claims (Rest, Narvaez, Bebeau & Thoma, in press). Instructional methods that promote development of the more adequate schemas for resolving ethical dilemmas have been shown to be effective in a wide range of educational settings (Schlaefli, Rest, & Thoma, 1985), including dentistry (Bebeau & Thoma, 1994), and performance on measures appears to predict to a wide range of prosocial behaviors, including clinical performance (Thoma, 1994). Yet, not all ethics courses consistently result in enhanced reasoning as measured by these tests

(e.g., Self & Baldwin, 1994; Self, Olivarez & Baldwin, 1994), nor are the effects consistently large. Such findings would argue for measures that are more closely targeted to the goals of the program.

Secondly, when researchers fail to find an effect on moral judgment for their ethics courses, some question the face validity of the test--arguing that the problems are not "current" problems and therefore are not a valid measure of their course. We would agree that the DIT may not be the best outcome measure for their purpose, but not because the test uses unfamiliar problems. Indeed, Westbrook (1994) showed that unfamiliar problems were better for assessing "bedrock concepts." In his study, journalists appeared to have learned responses to the familiar problems of their discipline without being able to apply the concepts to unfamiliar (or new) problems. Bedrock concepts are assumed to facilitate this transfer. A central question for any instructional program is whether instruction enables students to transfer learning to novel situations and whether "bedrock concepts" (the more general philosophical ideals or moral schemas) are a necessary condition for transfer of learning.

A third problem is described by Strike (1982). He argues that professionals who have a good grasp of abstract moral schemas may nevertheless lack an adequate grasp of specific moral concepts that apply to their profession. To prepare persons for life in a given profession, discipline-specific ethics courses must address specific concepts that have been worked out within the profession. Such specific concepts are thought to represent an intermediate level between the more general life-span schemas and the concrete prescription of ethical codes (Rest et al., in press). In health care professions, for example, intermediate concepts like professional autonomy, competence and autonomous choice, confidentiality, surrogate decision making, informed consent, privacy, veracity, fidelity, allocation or rationing of scarce resources, etc. need to be directly taught; whereas concepts like due process, whistle-blowing, intellectual freedom--central to other professions--are not emphasized as much in health professions. Intermediate level concepts are more concrete than general abstract principles, but even more concrete are professional codes of ethics. Codes concretely list specific prescriptions and prohibitions; they seldom provide a rationale or explanation from moral theory. Although students may need a working knowledge of the profession's code, professions differ in the emphasis placed on teaching and assessing knowledge of the professional code. Measures need to be designed that are specific to the concepts and goals of a given profession.

Strike's argument about the abstractness of Kohlberg's theory has an analogy in Piaget's theory. Consider a person who has reached Formal Operational thinking (a more general level of scientific thinking). Is acquisition of Formal Operational thinking sufficient to build a bridge or a computer (a specific level of scientific thinking)? In addition to Formal Operations, an engineer needs less abstract, intermediate level concepts. Course work in engineering addresses the intermediate concepts needed for successful practice in the discipline. Similarly, most professional ethics courses are organized around intermediate concepts rather than the more abstract general principles that may be taught in foundational courses in moral philosophy. Whether applied ethics courses enhance acquisition of the more general moral schemas, even when their focus is on more intermediate concepts, is a question of interest.

There are challenges in developing measures of intermediate level concepts. Some (e.g., Shalit, 1997) argue that there are no agreed-upon criteria for judging better and worse actions and justifications for moral problems, claiming that philosophers' recommendations in cases vary from culture to culture and from individual to individual. And, Fox and Stocking (1993) provide some empirical support for the difficulty in achieving consensus on action choices. They observed "little consensus" among the 154 ethicists who selected action choices for seven hypothetical cases involving a patient in a persistent vegetative state, cases the authors thought were simple and unambiguous. On the other hand, Rest and colleagues (Rest, Thoma, & Edwards, 1997) achieve consensus on better and worse justifications for moral problems presented on the Defining Issues Test (DIT), a test of life-span moral development. Perhaps it is only possible to achieve consensus

on the gross-grained distinctions reflected in items measuring life-span development (where researchers are attempting to distinguish the moral thinking of the 8 year old from that of 18 year old and 28 year old). Perhaps it is impossible to achieve acceptable levels of consensus on fine-grained professional ethical judgments. Design of a new measure must attend to issues of consensus and avoid the pitfall of searching for the single most appropriate response.

Defining construct validity for an intermediate concept measure presents an additional challenge. We do not conceptualize ethics expertise as a unitary construct that might be reflected in simple indices of internal consistency, but rather as a multifaceted performance dimension that draws upon expertise from several domains: technical knowledge, procedural knowledge, intermediate ethical concepts, and an understanding of ethical systems. Further, for an outcome measure of ethics expertise, educators would want to know whether a professional can discern the right course of action in problematic contexts, as well as to justify the action. Moral development theory and research (Rest, 1983) suggests that moral judgment and moral action are distinct and independent processes; thus, discerning among actions would not necessarily correlate with differentiating among justifications, especially for novices. Thus, our initial approach to measurement validation will focus on 1) strategies to insure the integrity of the cases and realism of the items for practicing dentists, 2) achieving consensus judgments among ethicists regarding better and worse actions and justifications, 3) exploring differences among groups expected to differ in dental ethics expertise, and 4) exploring relationships with other measures of expertise.

Specific Purpose

This study describes the steps taken to assure the integrity of the test, and reports the level of consensus achieved among dental ethicists as to better and worse action choices and justifications for the ethical problems presented. Using expert judgments to construct a scoring scheme, performance differences are reported for three groups expected to differ in dental ethics interest and expertise: college freshmen, entering dental students, and graduating senior dental students. Because the ICM was designed in the context of an ethics curriculum that has been in place for the past 15 years (Bebeau, 1994)--a curriculum that has consistently been shown to influence moral judgment development (Bebeau & Thoma, 1994), a cross-sectional comparison between senior dental students and entering freshmen enabled a first look at the educational effects of the ethics curriculum on intermediate concepts. Because the ICM included problems that were familiar to seniors, as well as problems that were novel, it was possible to examine the role that acquisition of more general philosophical ideals (i.e., the abstract moral schemas measured by a test of moral judgment) plays in a student's ability to select appropriate actions and justifications for unfamiliar problems.

Methods

Instrument development

Step 1. Five topics, representing some of the intermediate concepts taught during the curriculum, were selected for development of cases. To assure that the cases were realistic and sufficiently detailed to enable decision-making, cases were tested with dental faculty. Cases ranged from familiar to unfamiliar. One case (Suspicious Dentist) was exactly the same as a case presented in a third-year course. The next (Margaret Johnston) was parallel to a case presented in a fourth-year course. Technical information about the case was identical, but superficial details were changed (e.g., gender of the patient, occupation of the patient's parent, and so on). The third case (Clarkson) presented an issue (substandard work by a previously treating dentist) that had been presented in earlier cases. Technical information was different, as were the circumstances of the case, but the case could be resolved by applying the strategy suggested for one of the earlier cases. The fourth case (Block) presented an issue (professional vs. patient autonomy) that had been raised during instruction, but the case required application of the issue to a completely new situation. The last case (The Harry Johnson Case) presented an issue of professional monitoring, but the case presented a

rather subtle issue involving alleged emotional instability on the part of a referral specialist. Table 1 lists the concepts and action questions presented in each case.

Step 2 . The cases were presented in two sessions to 12 dental residents and 4 faculty, who were asked to study each case and to independently write out 1) what should be done and 2) why the proposed action choice was justified.

Step 3. The written responses were used to generate acceptable and unacceptable (though plausible from the perspective of some residents) action choice items and justification items for each case. In a third session, the 4 faculty and 12 residents again read each case, rated each action choice (highly defensible, defensible, questionable, not at all defensible) and then ranked the two best and two worst action choices. Respondents then rated the importance of each reason in coming to a decision (great, much, some, little, no) and ranked the three best and the two worst justifications.

Step 4. Based on responses, the action choices and justifications were modified and submitted to a second group (2 faculty and 3 advanced residents) who rated and ranked the actions and justifications, then discussed the cases and the responses in an effort to achieve consensus on the best and worst action choices and justifications. Again, modifications were made to clarify the cases and the items. A sample case is included in Appendix A.

Step 5. Eighteen teachers of dental ethics, who attended a meeting of the Professional Ethics in Dentistry Network (PEDNET) held in conjunction with the 1997 national meeting of the American Association of Dental Schools, were asked to respond to the cases. Fourteen completed the measure and returned it to the investigators in time to complete the analysis. A rank order of action choices and of justifications were developed for use as scoring guides. To assess expert agreement on the adequacy of the items, action choice and justification items were keyed by the developers as either appropriate, neutral, or inappropriate. Eighty-eight percent of the experts rated the action choice items keyed as appropriate as reasonable choices. Similarly, 88% of the experts rated the inappropriate actions as unreasonable action choices. For the justification data, the corresponding agreement was 95% for appropriate justifications and 93% for inappropriate justifications. Although there was good agreement on good and bad choices, and no miss classifications (i.e., rating a good choice as bad and visa versa) there was only moderate agreement as to what constituted the single best action or justification (range 32% to 46% agreement).

Scoring Procedures

To summarize ICM performance, four indices are computed for each case. Because it is possible for a respondent to rank an appropriate action or justification as inappropriate (or visa versa), negative numbers are possible. Indices for each case include: (1) the number of ranked items that matched expert consensus on appropriate action choices (potential range = -3 to +3) and (2) justifications (potential range = -6 to +6); (3) the number of ranked items that matched items considered inappropriate actions (potential range = -3 to +3) and (4) justifications (potential range = -3 to +3). Summed across cases, scores range from -15 to +15 for best and worst actions and worst justifications; -30 to +30 for best justifications. In order to compare performance across indices and across performance on familiar (two cases) and unfamiliar cases (three cases), scores are reported as the percent of actions or justifications that match judgments of the experts.

In addition to the four subscales, three summary indices were calculated: (1) an action choice total, combining best and worst actions; (2) a justification total, combining best and worst justifications; and (3) an ICM total, combining actions and justifications. To maintain consistency across indices and summary scores, these table values are also reported as percentages.

Data Collection

Samples. Three groups, representing three levels of dental ethics expertise, and three levels of moral judgment development, responded to the ICM. Group 1 consisted of 68 college freshmen attending a local liberal arts college. Students provided responses as part of a class project for an introductory psychology course. A mean DIT P scores for freshmen attending this liberal arts college is 33.4; for seniors 42.9 (McNeel, 1994). Groups 2 and 3 were 86 freshmen and 74 seniors from the University of Minnesota School of Dentistry. Mean DIT P scores for Minnesota dental freshmen (averaged over thirteen years) is 46.02; for seniors 51.13 (Bebeau & Thoma, 1994) All dental students completed The Defining Issues Test (DIT) and the newly devised test of intermediate ethical concepts (ICM) as part of either a baseline assessment (freshmen) or final assessment (seniors) for the school's ethics curriculum. Data for seniors were collected during spring quarter (1997) of their fourth year; data for freshmen were collected during the first week of the fall quarter (1997) of their first year; data for college freshmen were collected during the fall semester (1997). Dental school GPA was available for seniors as was their DIT P score as freshmen.

The Defining Issues Test (DIT). The DIT (Rest, 1979) is the most extensively validated and most widely used measure of moral judgment. The DIT assesses the proportion of time a person selects arguments representing Kohlbergian developmental frameworks when presented with six problems that represent a range of moral issues. Persons in professions, like medicine, dentistry, and law, select moral arguments that appeal to philosophical ideas (the more adequate conceptual frameworks--represented by the P score) about 50 percent of the time, whereas graduate students in moral philosophy select such arguments about 65 percent of the time (on average). Recent updates (Rest, Thoma, & Edwards, 1997; Rest, Thoma, Narvaez, & Bebeau, 1997) of the extensive literature on the validity and reliability of the measure confirms earlier reports (Rest, 1979).

Results

Construct validity

Group Differences on the ICM. Table 2 presents descriptive statistics for the three groups who competed the ICM. A one-way ANOVA with ICM summary scores (total ICM score; action choice total score; justification choice total score) as the dependent variables and group as the three-level, between-subjects factor indicate a significant main effect for Total ICM score ($F(2, 227) = 190.42$, $p < .00$); for action choice total score ($F(2,227) = 70.44$, $p < .00$); and for justification choice total score ($F(2, 227) = 49.57$, $p < .00$). Between group differences are in the expected direction (i.e., seniors score higher than dental freshmen who score higher than college freshmen) on each of the total score indices. Further, using the Tukey post hoc procedure, each between group comparison is statistically significant. In addition to the expected mean group differences, examination of the ranges reported in Table 2 indicates that miss classifications (i.e., selecting best actions and justifications as worst, or visa versa) are not found in the more expert group.

Correlations with other indices. For dental freshmen and seniors, findings indicate that the ICM total score was consistently and significantly related ($r = .26$ for freshmen; $r = .33$ for seniors) to DIT scores. However, none of the ICM indices were significantly related to GPA (Total score $r = .13$; action choice total $r = .10$; justification total $r = .04$). Such findings are consistent with the claim that the intermediate concept measure assesses ethical reasoning and judgment and not simple dental school performance.

Educational Intervention Effects

Expert/novice differences. Table 3 presents an in-depth comparison of "expert" and "less expert" dental students performance on each ICM index for familiar and unfamiliar cases. Consistent with an educational effect, the dental freshman to senior differences were evident across

the case studies and were larger in the familiar set of dilemmas. To test whether these observed differences are statistically reliable, four mixed-model ANOVAs were applied to the action choice (best/worst) and justification (best/worst) data. In each analysis, familiarity was the two-level within-subjects factor and class (i.e., freshman and senior) was the between-subjects factor. Of particular interest were the F statistics associated with class and the interaction term (i.e., the interaction between familiarity and class). The former F statistics indicates whether or not seniors performed better on the measure, regardless of whether or not the dilemmas were familiar. The interaction term, on the other hand, assesses the degree to which responses differed by class and familiarity. Thus if an expert-novice difference exists in the data, one would expect that seniors would do better than dental freshman overall, but especially on the familiar dilemmas.

Results of these analyses were consistent with expectations. Seniors were significantly higher than dental freshman on all four between-subjects comparisons ($F(1, 156) = 31.36$, $p < .01$; $F(1, 157) = 35.40$, $p < .01$; $F(1, 153) = 28.81$, $p < .01$; $F(1, 156) = 29.32$, $p < .01$, for identifying best action choice, best justification, worst action choice, and worst justifications, respectively). In addition, significant interaction terms were noted for best action choice selections ($F(1, 156) = 49.79$, $p < .01$) and best justification selections ($F(1, 157) = 8.06$, $p < .01$), but were not found for worst action choice ($F(1, 153) = 2.31$, $p = .13$) or worst justification selections ($F(1, 156) = 1.93$, $p = .17$). It appears therefore that educational effects are evident in the data, and are more strongly indicated for identifying appropriate action choices and justifications in familiar cases. On identifying unacceptable action choices and justifications, seniors were generally better than freshmen, however, the effect was not as clearly linked to prior exposure to the issues embedded within each case study.

Controlling for DIT P scores

It might be argued that intermediate concept measures offer little beyond that which can be obtained by more general moral schema measures such as the DIT. To assess the uniqueness of intermediate concept measures relative to DIT P scores, further analyses were conducted to estimate the degree to which DIT P score differences could account for the findings described above. In each of the four repeated measures ANOVAs, DIT P scores were treated as a covariate and the original findings reassessed. Results of these analyses suggest that P scores can not account for the findings described above. That is, all of the significant effects obtained in the earlier analyses remained statistically significant after controlling for DIT P scores. Thus, the expert-novice differences obtained on the intermediate concept measure could not be explained by DIT P scores.

Joint contribution of DIT P scores and training on the ICM

In spite of the fact that DIT P scores could not account for performance differences on the intermediate concept measure, it still may be that P scores, and the moral schema they represent, independently contribute to performance on intermediate concept measures. Particularly interesting is the potential joint contribution of training and moral schema on unfamiliar situations. It may be, for instance that when confronted with a novel situation, basic interpretive systems, such as moral schema, help to clarify the important ethical considerations and thus, facilitate the application of intermediate concepts in arriving at a solution.

Table 4 presents the summary statistics from hierarchical linear regression analyses that were designed to assess the unique contribution of DIT P scores on the intermediate concept measure indices after entering the expert-novice grouping. In each case, the intermediate concept index was regressed on the grouping factor (dummy coded 0 = Freshman; 1 = Seniors) and in a second step onto DIT P scores. Thus, the main goal of the analysis was to assess whether and to what degree DIT P scores contribute to intermediate concept scores after accounting for expert/novice effects. As expected from the previous analyses, the grouping factor accounted for a

substantial amount of the variance in intermediate concept measure scores assessed in familiar situations. In these same cases, DIT scores did not account for any additional variance in the intermediate concept measure scores.

Also expected was the relatively small amount of variance in intermediate concept scores on the unfamiliar case studies, due to expert novice differences. Consistent with the view that moral schema may complement intermediate concepts in unfamiliar situations, DIT P scores did account for additional variance on two of the four indices (i.e., identifying inappropriate action choices and appropriate justifications). A third finding reached a statistical tendency (i.e., identifying inappropriate justifications). Overall, therefore, there is evidence that moral schema, as measured by DIT scores, may contribute to performance on intermediate concept measures concerning unfamiliar situations.

Linking justifications and actions. One further expert-novice difference deserves mention. Consistent with the view that ethical training strengthens the connection between various aspects of ethical thinking (Taylor & Crocker, 1981), seniors were the only group to evidence strong relationships between justifications and action choices within both familiar ($r = .72$) and unfamiliar cases ($r = .75$), suggesting a clear link between the two processes. A similar relationship was not found for the novice groups on the familiar cases ($r = .27$). On unfamiliar cases the novice groups were more similar to the seniors ($r = .63$).

Discussion

Our findings support a continued focus on intermediate concepts as a new approach for understanding ethical reasoning. First, construct validity was demonstrated by dental ethicists who were able to achieve reasonable agreement on better and worse actions and justifications for the cases presented. This finding challenges the skepticism (Shalit, 1997; Fox & Stocking, 1993) about the potential for designing a profession-specific measure of ethics expertise. Even though the dental ethicists did not agree on a single best action or justification, agreement on appropriate and inappropriate actions and justifications was very high. There were no ratings that crossed the neutral barrier between appropriate and inappropriate items (i.e., no appropriate items rated as inappropriate, or visa versa). It appears that acceptable levels of consensus can be achieved on rather fine-grained measures of profession-specific ethics expertise.

Second, construct validity was also demonstrated by the fact that the exam discriminated among groups having different levels of dental ethics interest and education. Senior dental students scored significantly higher than dental freshmen, who in turn, scored significantly higher than college freshmen. Novice groups not only scored lower on the measure as a whole, but also miss classified items, ranking best choices as worst, and worst choices as best. Such findings are central to a claim of measurement validity for a test of ethics expertise.

Third, construct validity was demonstrated by the findings that DIT scores were consistently related to ICM scores. By contrast, GPA, a measure of academic aptitude and achievement was unrelated to ICM scores. Taken together, these data are consistent with the description of the ICM as a measure of ethical reasoning and not simply a measure of the dental curriculum. Further, the magnitude of the correlations with DIT scores suggests that the ICM is offering non-redundant information on the reasoning process.

Fourth, the intermediate concept measure was sensitive to advanced placement in the curriculum. Differences between freshman and senior dental students were highly significant and not synonymous with scores on a more broad-based measure of moral judgment development, nor were they related to dental school GPA. In other words, seniors were consistently better at resolving familiar and novel problems than freshmen, irrespective of a freshman's advanced standing on the measure of moral judgment development. Whereas seniors with high scores on the

measure of moral judgment had an edge over seniors with lower scores when attacking novel problems, ethics instruction appeared to account for the differences, rather than access to advanced moral schemas. The findings provide initial support for Strike's assessment, that moral judgment development alone will not enable professionals to deal with intermediate concepts. To prepare persons for life in a given profession, discipline-specific ethics courses must address specific concepts that have been worked out within the profession. This study suggests that such instruction makes a difference.

One possible explanation for the senior's poorer performance on unfamiliar ICM cases, is the possible differences in difficulty of the cases. There was no consistent pattern of differences in difficulty of the familiar vs. novel cases. Thus, it is not the case that poor performance by seniors on unfamiliar cases could be simply attributed to case difficulty. The findings do suggest different strategies for attending to intermediate-level concepts in cases that differ in familiarity. In familiar cases, seniors did not appear to rely on moral judgment schemas to help identify action choices and justifications, attempting instead to select actions and justifications consistent with the learned solutions to the case. Although appropriate action choices were readily identified using this strategy in familiar cases, participants were not able to better identify appropriate justifications and actions in unfamiliar situations. Thus, the skills learned in familiar cases did not necessarily translate to unfamiliar ones. By contrast, when a case is unfamiliar, individuals moral reasoning schemas did appear to be utilized in identifying and justifying action choices, and subjects with higher level conceptual schemas were better able to pick out the better action choices and justifications.

Taken together, these findings also point to the need for a multifaceted focus for a professional ethics curriculum. Although the data are cross-sectional, and longitudinal data are needed to strengthen claims for the utility of the measure, our data highlights the importance of presenting important ethical issues in detail, in order to provide students the means necessary to act in accordance with current expert views on the issues. Our data also support the need to attend to the more general schemas that individuals bring to bear on ethical issues that are novel and unexpected. Again, longitudinal data may help us pinpoint the role of general schemas in problem solving novel situations. Together, targeted instruction on central ethical issues, coupled with a clear focus on enhancing ethical reasoning, should prepare students for both expected and unexpected ethical dilemmas.

In summary, this study represents a first exploration of the usefulness of an intermediate concept measure. Although the cases used in this study may be used to assess expert novice differences in other dental school settings, the cases do not necessarily represent the full range of issues or circumstances dentists may encounter in practice. Whereas the method of assessment shows initial promise, several steps are required to design a measure that would be useful across dental education settings. Before proceeding with further case development, it may be necessary to first survey the discipline to determine: (1) the concepts currently taught, (2) the concepts ethicists and practitioners in the discipline think are essential, and (3) the kinds of problems that frequently occur. Several studies of this nature have already been conducted in dentistry, and it may be possible to work from existing data bases to more quickly achieve consensus on concepts and cases. Working together, the dental ethicists that form the Professional Ethics in Dentistry Network (PEDNET) could make this an ongoing project that would strengthen professional ethics education. Grounding the design of the measure on a data-base, generated by a wide range of practitioners and ethicists in the discipline, would increase the utility and validity of the measure.

References

American Association of Dental Schools. (1989). Curriculum guidelines on ethics and professionalism in dentistry. Journal of Dental Education, 53(2), 144-148.

American Dental Association. Commission on Dental Accreditation. (1991). Accreditation standards for dental education programs. Chicago: American Dental Association.

Bebeau, M. J. (1994). Can ethics be taught? A look at the evidence: Revisited. New York State Dental Journal, 60(1), 51-57.

Bebeau, M. J. (1994). Influencing the moral dimensions of dental practice. In J. R. Rest, & Narvaez, D. F. (Eds.), Moral development in the professions: Psychology and applied ethics (pp. 189-198). Hillsdale, NJ: Erlbaum.

Bebeau, M. J., & Thoma, S. J. (1994). The impact of a dental ethics curriculum on moral reasoning. Journal of Dental Education, 58(9), 684-692.

Colby, A., & Kohlberg, L. (1987). The measurement of moral judgment: Volumes I and II. Cambridge MA: Cambridge University Press.

Fox, E., & Stocking, C. (1993). Ethics consultants' recommendations for life-prolonging treatment of patients in a persistent vegetative state. Journal of the American Medical Association, 270(21), 2578-2582.

Gibbs, J. C., Basinger, K. S., & Fuller, D. (1992). Moral maturity: Measuring the development of sociomoral reflection. Hillsdale, NJ: Erlbaum.

Hastings Center. (1980). The teaching of ethics in higher education. Hastings-on-Hudson, NY.

Lancaster, D. M., Gardiner, J. F., Strother, E. A., & Boozer, C. H. (1989). Dental students' attitudes: A four year study. Journal of the American College of Dentists, 56(4), 30-35.

McNeel, S. P. (1994). College teaching and student moral development. In J. R. Rest, & D. F. Narvaez (Eds.), Moral development in the professions: Psychology and applied ethics (pp. 27-50). Hillsdale, NJ: Erlbaum.

Moody, P. M., Van Tassel, C., & Cash, D. M. (1974). Cynicism, humanitarianism, and dental career development. Journal of Dental Education 38(11), 645-649.

Morris, R. T., & Sherlock, B. J. (1971). Decline of ethics and the rise of cynicism in dental school. Journal of Health and Social Behavior, 12(4), 290-299.

Parrish, J. R. (1968). Professional conduct in dental school and after. Journal of Dental Education, 32(3), 326-329.

Rest, J. R. (1979). Development in judging moral issues. Minneapolis: University of Minnesota Press.

Rest, J. R., (1983). Morality. In P. H. Mussen (Series Ed.) & J. Flavell & E. Markman (Vol. Ed.), Handbook of child psychology: Vol. 3, Cognitive Development (4th ed., pp. 556-629). New York: Wiley.

Rest, J. R., Narvaez, D., Bebeau, M. J., & Thoma, S. J. (in press). Moral judgment development in adolescents and adults: A neo-Kohlbergian approach based on the Defining Issues Test. Hillsdale, NJ: Erlbaum.

Rest, J. , Thoma, S., & Edwards, L. (1997). Designing and validating a measure of moral judgment: Stage preference and stage consistency approaches. Journal of Educational Psychology, 89(1), 5-28.

Schlaefli, A., Rest, J. R., & Thoma, S. J. (1985). Does moral education improve moral judgment? A meta-analysis of intervention studies using the Defining Issues Test. Review of Educational Research, 55(3), 319-352.

Self, D. J., & Baldwin, D. C. (1994). Moral reasoning in medicine. In J. R. Rest, & D. F. Narvaez (Eds.), Moral development in the professions: Psychology and applied ethics (pp. 147-162). Hillsdale, NJ: Erlbaum.

Self, D. J., Olivarez, M., & Baldwin, D. C. (1994). Moral reasoning in veterinary medicine. In J. R. Rest, & D. F. Narvaez (Eds.), Moral development in the professions: Psychology and applied ethics (pp. 163-172). Hillsdale, NJ: Erlbaum.

Shalit, R. (1997, April 28). When we were philosophers' kings. The New Republic, 24-28.

Steinberg, D. N. (1973). Change in attitudes of dental students during their professional education. Journal of Dental Education, 37(5), 36-41.

Strike, K. A. (1982). Educational policy and the just society. Chicago: University of Illinois Press.

Taylor, S. E. & Crocker, J. (1981) Schematic bases of social information processing. In E. T. Higgins, & C. P. Herman & M. P. Zanna (Eds.) Social cognition: The Ontario Symposium. Vol. 1 (pp. 89-134). Hillsdale, NJ: Erlbaum.

Thoma, S. J. (1994). Moral judgment and moral action. In J. R. Rest, & D. F. Narvaez (Eds.), Moral development in the professions: Psychology and applied ethics (pp. 199-212). Hillsdale, NJ: Erlbaum.

Thornton, B. C., Callahan, D., & Nelson, J. L. (1993). Bioethics education: Expanding the circle of participants. Hastings Center Report, 23(1), 25-29.

Westbrook, T. (1994). Tracking the moral development of journalists: A look at them and their work. In J. R. Rest, & D. F. Narvaez (Eds.), Moral development in the professions: Psychology and applied ethics (pp. 189-198). Hillsdale, NJ: Erlbaum.

Table 1
Intermediate Concepts and Action Questions Presented in Each Case Presented on Test

Case	Intermediate Concept	Action Question
Clarkson	veracity/ professional monitoring	What to do when a colleague's care fails to meet professional standards
John Block	paternalism/ patient autonomy	What to do when the patient requests treatment the dentist considers harmful
Harry Johnson	fidelity/ professional monitoring	What to do when you hear rumors about the emotional stability of a specialist to whom you refer patients
Margaret Johnston	conflicting commitments	What to do when a child appears to be abused and the parent appears to be misrepresenting what happened
Suspicious Dentist	paternalism/ patient autonomy	What to do when a patient requests a controlled substance for alleged pain

Table 2
Group Differences on ICM Scores and DIT P Scores

	Groups											
	A College Freshmen[a]			B Dental Freshmen[b]			C Dental Seniors[c]			Between Group Effect Size[d]		
	M	SD	Range	M	SD	Range	M	SD	Range	AB	BC	AC
ICM scores												
Action Choice	21.32	18.45	-27 to 63	39.96	17.55	-17 to 80	55.72	15.72	23 to 90	.93	.85	1.42
Justification	30.42	15.39	-13 to 58	38.86	14.05	4 to 69	53.06	11.67	18 to 76	.56	.96	1.28
ICM Total	26.78	13.13	-13 to 52	39.30	12.69	-4 to 65	54.13	10.09	24 to 75	.88	1.08	1.52
DIT P scores				46.54	12.64		48.63	13.46			.16	

Note. The values represent mean percent of action or justification choices that agree with judgments derived from responses of 14 dental ethicists.

[a] $n = 68$. [b] $n = 82$. [c] $n = 73$. [d] Effect size was determined by subtracting the "less expert" group average from "more expert" group average and dividing by the pooled-group standard deviation.

Table 3
Differences Between Expert and Novice Dental Students on Familiar and Unfamiliar ICM Cases

	Dental Freshmen[a]		Dental Seniors[b]		
	M	SD	M	SD	Effect Size[c]
ICM scores					
Action choice					
Familiar					
Best	43.80	27.89	86.26	20.71	1.71
Worst	50.19	28.87	71.46	20.50	.78
Unfamiliar					
Best	42.61	36.73	40.33	30.34	-.07
Worst	28.86	30.13	40.84	30.17	.39
Justifications					
Familiar					
Best	29.65	26.28	51.58	20.26	.85
Worst	13.33	23.96	32.21	20.33	.79
Unfamiliar					
Best	54.38	19.17	63.36	18.27	.46
Worst	38.11	29.69	48.86	21.26	.41

Note. The values represent mean percent of action or justification choices that agree with judgments derived from responses of 14 dental ethicists.
[a] $n = 82$. [b] $n = 73$. [c] Effect size was determined by subtracting the freshmen from the senior average and dividing by the pooled-group standard deviation.

Table 4
Joint Contribution of Expert-Novice (Group) and Moral Schema (DIT P) on Familiar and Unfamiliar ICM Cases

	Familiar		Unfamiliar	
	Group	DIT P	Group	DIT P
Action choice				
Best	.42*	.02	.00	.01
Worst	.39*	.00	.04*	.06*
Justifications				
Best	.43*	.02	.06*	.05*
Worst	.40*	.01	.04*	.03†

Note. Numbers represent R2 values. In each case, group was entered first in the hierarchy followed by DIT P scores. Thus, the numbers associated with DIT P scores represents the amount of additional variance in the intermediate concept measure attributed to these scores.
*$p < .05$; †$p < .1$.

Appendix A

Intermediate Concepts Measure

Test Directions:

Read each case. Then, rate and rank the alternative action choices. When you have finished, turn the page and rate and rank the reasons to justify your action choices.

Sample Case

The John Block Case

Dr. Meyer looked at his schedule and noticed that his next patient for the day was John Block. He wondered what brought John in today, as it hadn't been too long since John's annual check-up. John had been a patient for several years, and was quite health conscious. He worked out regularly at the local health club. He took reasonably good care of his teeth, and kept his regular hygiene appointments. When Dr. Meyer asked what brought John to see him today, John said that he wanted his lower second molar extracted. Dr. Meyer looked at the chart and noticed that he had performed root canal therapy (using a standard procedure) and placed a crown on that tooth two years ago. It had been John's first and only experience with root canal therapy. John had been somewhat apprehensive at the time, but the treatment had been uneventful, and the crown appeared to be in good condition.

When Dr. Meyer asked if the tooth was bothering him, John said it wasn't, but wanted the tooth removed nonetheless. When Dr. Meyer asked why, John handed him a brochure from the health club that included information about the toxic nature of material used in root canal therapy and advised against the procedure. Dr. Meyer said he didn't believe the information in the newsletter was true, but John was adamant. He said that this wasn't the only place he had read about this, and he wanted foreign material out of his mouth. "I'm reevaluating my health," said John, "and I'm making a number of changes in my health habits, exercise routines, and diet to improve my health. I'm even becoming more meticulous about my oral hygiene." When Dr. Meyer inquired if he had any particular health problems, John indicated that he was simply trying to lead a healthier lifestyle. After several more minutes of conversation, Dr. Meyer came to the conclusion that further discussion was fruitless. John asserted that it was his tooth after all; he had paid for the treatment and he now wanted the tooth to be removed. He thought the dentist should honor his request. Further, when the dentist had given him his options about whether or not to do the root canal, pulling the tooth had been one of the options presented. He failed to see how Dr. Meyer could fail to comply with his request now. Was Dr. Meyer one of those dentists who does what he wants, rather than what the patient wants?

The John Block Case: Action Choices

Dr. Meyer thought about the clinical ambiguities and considered several alternatives. Rate each action choice using the following scale:

HD = Highly Defensible D = Defensible Q = Questionable ND = Not at all Defensible

HD D Q ND 1. Suggest that John see another dentist for another opinion.

HD D Q ND 2. Refuse to pull the tooth.

HD D Q ND 3. Try to further educate the patient.

HD D Q ND 4. Respect the patient's wishes and pull the tooth.

HD D Q ND 5. Have the patient sign a consent form indicating that he has been informed of the consequences of pulling the tooth. Then, pull the tooth.

HD D Q ND 6. Tell the patient that you respect his right to do whatever he chooses, but you won't pull the tooth.

HD D Q ND 7. Tell the patient that you took an oath not to cause harm and removing the tooth would do more harm than good.

Pick the two best action choices:

Best Choice _____
Second Best _____

Pick the two worst action choices:

Worst Alternative _____
Next Worst _____

On the next page, you are asked to consider the reasons that best justify your choice.

The John Block Case: Justification for Action Choices

Rate the importance of each reason, in coming to your decision, by circling your choice.

G = Great M = Much S = Some L = Little N = No

G M S L N 1. Dr. Meyer's colleagues will not approve of the removal of the tooth.

G M S L N 2. Doing work that is not required cannot be justified.

G M S L N 3. The dentist is responsible for judging the scientific merit of a treatment decision.

G M S L N 4. The dentist shouldn't let the patient control the treatment decisions.

G M S L N 5. If you don't do what the patient wants, you are likely to lose business.

G M S L N 6. The patient's right to decide what will happen to their bodies needs to be respected.

G M S L N 7. The dentist needs to be open-minded about alternative therapies.

G M S L N 8. You need to be respectful of the patient's viewpoints when asserting your professional judgment.

G M S L N 9. If you wouldn't remove a patient's finger because he felt it lead to a healthier life style, you shouldn't take out a tooth because of a patient's belief that it lead to a healthier lifestyle.

G M S L N 10. If the patient is adamant about a decision, and has been properly educated and warned of the consequences, the dentist should do what the patient wants.

G M S L N 11. If the patient wants something done that is not a big health risk, then why not do it?

G M S L N 12. In the long run, it's better to give up a little authority than to lose this patient from your practice.

From the list above, pick the three best reasons: Next, pick the two worst reasons:

Most Important Reason _____
Second Most Important _____
Third Most Important _____

Worst Reason _____
Next Worst _____